REMARKS

Claims 1–21 are pending in the application.

Claims 1–21 have been rejected.

Claims 1–3, 5–8, 13–14 and 18 have been amended, as set forth herein. Claims 3, 5 and 7 were amended solely to place those claims in independent form, including all limitations of the base claim and any intervening claims, without altering the scope of those claims. Claim 8 was amended solely to correct antecedent basis and typographical errors therein, without altering the scope of that claim. Claims 13–14 and 18 were amended solely to correct antecedent basis, punctuation and grammar errors therein, without altering the scope of those claims.

Reconsideration of the claims is respectfully requested.

I. REJECTION UNDER 35 U.S.C. § 102

Claims 2, 6, 10 and 12–13 were rejected under 35 U.S.C. § 102(b) as being anticipated by U.S. Patent No. 5,719,921 to *Vysotsky et al.* The rejection is respectfully traversed.

A cited prior art reference anticipates the claimed invention under 35 U.S.C. § 102 only if every element of a claimed invention is identically shown in that single reference, arranged as they are in the claims. MPEP § 2131; *In re Bond*, 910 F.2d 831, 832, 15 U.S.P.Q.2d 1566, 1567 (Fed. Cir. 1990). Anticipation is only shown where each and every limitation of the claimed invention is found in a single cited prior art reference. MPEP § 2131; *In re Donohue*, 766 F.2d 531, 534, 226 U.S.P.Q. 619, 621 (Fed. Cir. 1985).

Independent claims 2 and 6 each recite that the first set of (user-independent) word models

includes models for each of a plurality of words, and the second set of (speaker dependent) word models includes models for at least some of that plurality of words. Similarly, independent claims 10 and 13 each recite that the first set of (user independent) recognition models relates to a plurality of system commands, and the second set of (user specific) recognition models relates to at least some of those system commands. Thus, independent claims 2, 6, 10 and 13 each include some overlap in the sets of words or commands modeled by user-independent and user-specific word/recognition models. Such a feature is not described by the cited reference. *Vysotsky et al* depicts and recites a system speaker dependent speech templates for recognizing names (within the user's telephone number directory) and speaker independent speech templates for commands (such as "Call Return"). *Vysotsky et al*, Figure 4, elements 403–404, 408 and 410; Abstract; column 11, lines 15–33. *Vysotsky et al* is silent as to employing both speaker dependent and speaker independent templates for overlapping sets of names or commands.

As noted, independent claims 10 and 13 each recite that the first set of (user independent) recognition models relates to a plurality of system commands, and the second set of (user specific) recognition models relates to at least some of those system commands. Such a feature is not described by the cited reference. *Vysotsky et al* recites speaker independent speech templates for commands, but only describes speaker dependent speech templates for names. *Vysotsky et al* is silent as to speaker dependent speech templates for at least some commands.

Claim 12 recites, for each model within the second (user specific) set and prior to storing that model, comparing the user specific recognition model to each of the first set of (user independent)

recognition models and other existing models within the second set, to ensure that speech recognized using each user specific recognition model will not be mistakenly recognized using any model in the first set or other models in the second set. Such a feature is not described in the cited reference. The cited portion of *Vysotsky et al* recites determining whether a name is too similar to another name in the customer's directory to reduce the risk of recognition errors:

> In accordance with an exemplary embodiment of the present invention, when a user is in a directory maintenance mode of operation and a speaker dependent model for a name to be added to the customer's directory is being generated, two repetitions of the name spoken by the customer are required. The system of the present invention seeks a consistent pair of utterances before building a model, and additionally determines if the name is too similar to another name in the customer's directory to reduce the risk that the name will be miss-recognized [sic] due to the use of multiple similar names. In order to achieve this, a recognition pass is performed during the training process to asses [sic] the system's ability to properly detect the name being added to the directory.

Vysotsky et al, column 8, lines 54–67. However, as noted above, *Vysotsky et al* recites only speaker independent (first set) speech templates for commands. *Vysotsky et al* is silent as to comparing speaker dependent (second set) speech templates for names to such speaker independent speech templates for commands.

Accordingly, the Applicant respectfully requests the Examiner withdraw the § 102(b) rejection of Claims 2, 6, 10 and 12–13.

II. REJECTION UNDER 35 U.S.C. § 103

Claims 1, 11 and 14–21 were rejected under 35 U.S.C. § 103 as being unpatentable over *Vysotsky et al* in view of U.S. Patent No. 5,377,303 to *Firman*. Claims 3–4 and 7–9 were rejected under 35 U.S.C. § 103 as being unpatentable over *Vysotsky et al* in view of U.S. Patent No.

4,618,984 to *Das et al.* Claim 5 was rejected under 35 U.S.C. § 103 as being unpatentable over *Vysotsky et al* in view of U.S. Patent No. 5,774,841 to *Salazar et al.* These rejections are respectfully traversed.

In *ex parte* examination of patent applications, the Patent Office bears the burden of establishing a *prima facie* case of obviousness. MPEP § 2142; *In re Fritch*, 972 F.2d 1260, 1262, 23 U.S.P.Q.2d 1780, 1783 (Fed. Cir. 1992). The initial burden of establishing a *prima facie* basis to deny patentability to a claimed invention is always upon the Patent Office. MPEP § 2142; *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992); *In re Piasecki*, 745 F.2d 1468, 1472, 223 U.S.P.Q. 785, 788 (Fed. Cir. 1984). Only when a *prima facie* case of obviousness is established does the burden shift to the applicant to produce evidence of nonobviousness. MPEP § 2142; *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992); *In re Rijckaert*, 9 F.3d 1531, 1532, 28 U.S.P.Q.2d 1955, 1956 (Fed. Cir. 1993). If the Patent Office does not produce a *prima facie* case of unpatentability, then without more the applicant is entitled to grant of a patent. *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992); *In re Grabiak*, 769 F.2d 729, 733, 226 U.S.P.Q. 870, 873 (Fed. Cir. 1985).

A *prima facie* case of obviousness is established when the teachings of the prior art itself suggest the claimed subject matter to a person of ordinary skill in the art. *In re Bell*, 991 F.2d 781, 783, 26 U.S.P.Q.2d 1529, 1531 (Fed. Cir. 1993). To establish a *prima facie* case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to

modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations. The teaching or suggestion to make the claimed invention and the reasonable expectation of success must both be found in the prior art, and not based on applicant's disclosure. MPEP § 2142.

Like independent claims 2 and 6 discussed above, independent claim 1 recites that the first set of (user-independent) word models includes models for each of a plurality of words, and the second set of (speaker dependent) word models includes models for at least some of that plurality of words. As noted, *Vysotsky et al* is silent as to such a feature. The cited portion of *Firman* is also silent as to such a feature.

Independent claims 3 and 7 each recite inviting a particular user, upon first use of the speech recognition system, to speak training words for deriving the second set of (speaker dependent) word models. Such a feature is not described in the cited reference. As conceded in the Office Action, *Vysotsky et al* does not describe such a feature. The cited portion(s) of *Das et al* only recites prompting the user to provide a new prototype utterance only for unrecognized utterances:

> Once the utterance choice number is determined, as identified by "YES" pathway 31 to PROMPT CHOICE box 32, the user is provided with a prompt requesting that the user make the utterance of choice. This utterance is the utterance to be used in certain situations as the new prototype.
>
> . . . .
>
> FIG. 2 details the procedures for adaptive retraining of prototype vocabulary items. During the recognition process if the talker produces a keyword mistake the adaptive retraining routine is invoked. Under computer control the procedures are entered and a stack of vocabulary item choices--related to the words just before the keyword mistake recognition--is sent to the adaptive training stage. Using this stack

> the system prompts the user to indicate which stack vocabulary item was uttered just prior to its identification as an unrecognized utterance. (This prompting can either be by audio response or character display from a gas panel, cathode ray tube, or the like.) If the word equivalent to the misrecognized utterance is not contained in the stack, the user is prompted to recite the utterance again, and the recognition procedure continues. Upon acknowledgment by the talker that a stack vocabulary item matches the utterance (INPUT 1), the user is prompted to provide a new prototype keyword candidate (INPUT 2) appropriate to that vocabulary item. The system calculates the matrix distances [INPUT 1 vs. INPUT 2=(D1)], [PROTOTYPE vs. INPUT 2=(D2)], and [PROTOTYPE vs. INPUT 1=(D3)], where INPUT 1 is the misrecognized utterance and INPUT 2 is the prompted utterance of the same keyword. If D1 is less than D2, and if D1 is less than D3, then INPUT 2 becomes the new prototype keyword. Otherwise the old prototype is retained. This procedure guarantees the best currently available prototype relative to the current user's (the user's current) speech characteristics. This procedure is independent of the mode of determining the particular matrix distance characteristics employed in the recognition process.

Das et al, column 4, lines 55–60 and column 5, line 45 through column 6, line 9. *Das et al* does not describe inviting a new user to speak a set of training words to derive user-specific prototypes. In fact, *Das et al* specifically states that such training for a new user is NOT required once a single initial prototype is acquired:

> The invention provides an optimum technique for prototype establishment involving merely an initial single prototype statement of each vocabulary item (from the first talker or electronic equivalent) and thereafter requiring no recitation of vocabulary items by either the first or subsequent talkers--except for those vocabulary items in which the system has difficulty performing correct recognition.

Das et al, column 2, lines 56–63.

Claims 4 and 8 each recite inviting the user to speak training utterances for deriving a user-dependent word model upon a predetermined number of failures to recognize the word using user-independent word models. Such a feature is not described in the cited references. As conceded in

the Office Action, *Vysotsky et al* is silent as to such a feature. Furthermore, as set forth above, *Das et al* recites prompting the user to enter a new prototype each time an utterance is not recognized and the user confirms that the utterance was of a prototype word. *Das et al*, column 2, lines 56–63. However, *Das et al* is silent at to prompting the user for a new prototype only after a predetermined number of recognition failures.

Claim 5 recites determining a likelihood of recognizing a spoken word using a first set of (speaker independent) word models, and deriving a word model from a spoken word marginally recognized using the first set for storage within the second set of (speaker dependent) word models. Such a feature is not described in the cited references. As conceded in the Office Action, *Vysotsky et al* does not describe using a marginally recognized spoken word to derive a word model. The cited portion of *Salazar et al* recites dynamically adapting recognition acceptance and delta thresholds for a word for which a recognition score indicates a low degree of confidence, to capture time-of-day, fatigue and illness-related variances:

> If, however, either of the two above criteria are not met, the word spoken has a low degree of confidence. In that case, the processor 130 first instructs the speech recognizer board 160, by means of instructions sent over the bus 120 into the interface 163, to send the query buzzer tone back to the user by means of the communication path 161b, through headset and VTR interface 141, and subsequently out on the communication path 52 to the headset 20, and simultaneously out communication path 57 to a VTR It should be pointed out that under software control, the processor can change the recognition acceptanc [sic] and delta threshold parameters dynamically. This is advantageous when poor recognition performance is occurring even when microphone calibration and confidence check has been done. By relaxing these parameters, some better recognition performance can be obtained. Next, the computer 130 uses only the first two words returned from the speech recognizer board 160 for the query. The processor 162 then sends one word at a time to the display 10 along with a "yes/no" verbal response. If a "yes" is recognized for

> the word on the display, the processor instructs the speech recognizer to adapt the word. At that point, the word adaptation update is placed in random access memory (RAM) of the speech recognizer circuit 160. It is not until the ASRU is placed in standby, either by voice command or manually by switch, that the updated vocabulary is stored in flash EPROM in the computer 130 to ensure the updated vocabulary is permanently stored. Query occurs during the actual operation of the ASRS when two questionable recognitions occur consecutively. That is, the ASRU can query the user at any time the ASRS is in the operation mode. Query allows capturing the actual annunciation of the spoken commands during the use of the ASRS in the actual application and at a given time of day. The user's annunciation of words over different periods of time is captured, which may therefore include annunciation when the user is fatigued or sick.

Salazar et al, column 14, line 44 through column 15, line 17. However, *Salazar et al* only recites such adaptation in the context of user-specific recognition criteria. *Salazar et al* is silent as to deriving a user-dependent word model from a spoken word marginally recognized using a user-independent word model.

Claim 8 recites that the user is invited to speak training sequences when a word (which could not be recognized within a predetermined number of attempts) does not have a model present in the user-dependent word models. Such a feature is not described by the cited references. As conceded in the Office Action, *Vysotsky et al* is silent as to this feature. Moreover, *Das et al* recites prompting the user to speak a new prototype only when the utterance corresponds to a stack entry. *Das et al*, column 2, lines 56–63. *Das et al* is silent as to inviting the user to speak a training sequence when no user-dependent word model exists for a particular word.

Claims 15 and 20 each recite prompting a user to record utterances to take the place of system command(s), allowing the user to employ a single "shorthand" word for a series of commands and/or avoiding or minimizing the potential for error between user-dependent and user-independent words

models by having the user-dependent word model relate to a different word actuating the same command as a user-independent word model (e.g., "dial" instead of "call" for the user-dependent word model to actuate the call command). Such a feature is not described in the cited references. *Vysotsky et al* is silent as to prompting the user to record utterances in place of system commands. *Firman* recites allowing the utterance for a voice control to differ from the command name for that voice control:

> In this mode, the Recognizer Software 120 displays to the user a menu of the utterance names (such as "file", "page down") which are to be recognized. These names, and the corresponding Voice Control command strings (indicating the appropriate actions) appear in a current word list 124. The user designates the utterance name of interest and then is prompted to speak the utterance corresponding to that name. For example, if the utterance name is "file" the user might utter "FILE" or "PLEASE FILE". The digitized samples from the Voice Navigator box 112 corresponding to that utterance are then used by the Recognizer Software 120 to create a "macro" representing the utterance, which is stored in the voice file 122 and subsequently associated with the utterance name in the word list 124. Ordinarily, the utterance is repeated more than once, in order to create a macro for the utterance that accommodates variation in a particular speaker's voice.
>
> The meaning of the spoken utterance need not correspond to the utterance name, and the text of the utterance name need not correspond to the Voice Control command strings stored in the word list. For example, the user may wish a command string that causes the operating system to save a file to have the utterance name "save file"; the associated command string may be "@MENU(file,2)"; and the utterance that the user trains for this utterance name may be the spoken phrase "immortalize". The Recognizer Software and Voice Control cause that utterance, name, and command string to be properly associated in the voice file and word list 124.

Firman, column 8, lines 25–55. However, while permitting different utterances and command names, the cited portion of *Firman* is silent as to prompting (or otherwise requiring) the user to record an utterance different from the command name, to be employed in place of the command name. The Office Action does not identify a description of this feature anywhere in (the 557 pages

of) *Firman*.

Like claim 12, claims 16 and 21 each recite, for each model within the second (user specific) set and prior to storing that model, comparing the user specific recognition model to each of the first set of (user independent) recognition models and other existing models within the second set, to ensure that speech recognized using each user specific recognition model will not be mistakenly recognized using any model in the first set or other models in the second set. Such a feature is not described in the cited references. *Vysotsky et al*, as noted above, does not describe this feature. Moreover, the Office Action does not identify a description of this feature anywhere in *Firman*. The cited portion of *Firman* does not describe such a feature.

Like independent claims 10 and 13, independent claim 18 recites that the first set of (user independent) recognition models relates to a plurality of system commands, and the second set of (user specific) recognition models relates to at least some of those system commands. Such a feature is not described by the cited references. *Vysotsky et al*, as noted above, does not describe such a feature. The cited portion of *Firman* does not describe such a feature.

Accordingly, the Applicant respectfully requests withdrawal of the § 103 rejection of Claims 1, 3–5, 7–9, 11 and 14–21.

III. REQUIREMENT UNDER 35 U.S.C. § 132/37 C.F.R. § 1.78(c)

The Office Action contains a requirement pursuant to 35 U.S.C. § 132 and 37 C.F.R. § 1.78(c) of a showing that the conflicting inventions were commonly owned at the time the invention was made of to name the prior inventor of the conflicting subject matter.

Pursuant to 10 C.F.R. § 10.18, following a reasonable inquiry, the undersigned certifies that the above-identified application and U.S. Patent No. 6,487,530 issued on application serial no. 09/281,078 to *Lin et al* were, at the time the invention claimed in the above-identified application was made, owned by the same person and/or subject to an obligation of assignment to the same person, NORTEL NETWORKS LIMITED.

IV. DOUBLE PATENTING

Claims 1–4 and 6–7 were rejected under the judicially created doctrine of obviousness-type double patenting as being directed to an invention not patentably distinct from claims 1–2, 4 and 6–7 of application serial no. 09/281,078, now U.S. Patent No. 6,487,530. This rejection is respectfully traversed.

A double patenting rejection of the obvious-type is "analogous to [a failure to meet] the nonobviousness requirement of 35 U.S.C. 103" except that the patent principally underlying the double patenting rejection is not considered prior art. MPEP § 804.02(II)(B)(1); *In re Braithwaite*, 379 F.2d 594, 154 U.S.P.Q. 29 (CCPA 1967). Therefore any analysis employed in an obvious-type double patenting rejection parallels the guidelines for analysis of a 35 U.S.C. 103 obviousness determination. MPEP; 804.02(II)(B)(1); *In re Braat*, 937 F.2d 589, 19 U.S.P.Q.2d 1289 (Fed. Cir. 1991); *In re Longi*, 759 F.2d 887, 225 U.S.P.Q. 645 (Fed. Cir. 1985).

Independent claim 1 recites that the first set of (user-independent) word models includes models for each of a plurality of words, and the second set of (speaker dependent) word models includes models for at least some of that plurality of words. Similarly, independent claims 2 and 6

each recite that the first set of (user-independent) word models includes models for each of a plurality of words, and the second set of (speaker dependent) word models includes models for at least some of that plurality of words. Such a feature is not recited in claims 1–2, 4 and 6–7 of U.S. Patent No. 6,487,530.

Accordingly, the Applicant respectfully requests withdrawal of the double patenting rejection of Claims 1–2 and 6. With respect to claims 3–4 and 7 of the subject application, Applicant respectfully requests that the requirement of a terminal disclaimer be deferred until those claims have been determined to be otherwise allowable.

V. CONCLUSION

As a result of the foregoing, the Applicant asserts that the remaining Claims in the Application are in condition for allowance, and respectfully requests an early allowance of such Claims.

AMENDMENTS WITH MARKINGS TO SHOW CHANGES MADE

Claims 1–3, 5–8, 13–14 and 18 were amended herein as follows:

1. (amended) A speech recognition system comprising computer memory storing:

a first set of speaker-independent word models used to match a word in an utterance of a user with a word model in said first set, wherein said first set of word models includes models for each of a plurality of words;

a second set of speaker dependent word models derived from speech of a particular user and used to match a word in an utterance of said particular [speaker] user, wherein said second set of word models includes models for at least some of said plurality of words; and

a program portion used to identify words in utterances of said particular user by attempting to match portions of an audio signal with:

word models among said first set; and

word models among said second set.

2. (amended) A method of operating a speech recognition system recognition system comprising:

storing a first set of speaker-independent word models used to match a word in an utterance of any user with a word model in said first set, said first set of word models including models for each of a plurality of words;

storing a second set of speaker dependent word models derived from speech of a particular user, said second set of word models including models for at least some of said plurality of words; and

recognizing words in utterances of said particular user by attempting to match portions of an audio signal with:

word models among said first set; and

word models among said second set.

3. (amended) [The] A method [according to claim 2 further] of operating a speech recognition system recognition system comprising:

storing a first set of speaker-independent word models used to match a word in an utterance of any user with a word model in said first set;

storing a second set of speaker dependent word models derived from speech of a particular user by:

inviting said particular user upon first use of said speech recognition system to speak training words for deriving said second set;

deriving said second set from said training words; and

storing said second set; and

recognizing words in utterances of said particular user by attempting to match portions of an audio signal with:

word models among said first set; and

word models among said second set.

5. (amended) [The] A method [according to claim 2 further] of operating a speech recognition system recognition system comprising:

storing a first set of speaker-independent word models used to match a word in an utterance of any user with a word model in said first set;

storing a second set of speaker dependent word models derived from speech of a particular user by:

determining a likelihood of recognizing a spoken word using said first set;

deriving a word model from a spoken word marginally recognized using said first set;

storing said word model in said second set; and

recognizing words in utterances of said particular user by attempting to match portions of an audio signal with:

word models among said first set; and

word models among said second set.

6. (amended) A method of enhancing speech recognition comprising:

providing a set of user-independent word models derived from utterances of a plurality of speakers, said first set of word models including models for each of a plurality of words;

providing a set of user-dependent word models for ones of a plurality of users each derived from utterances of one of said users, said second set of word models including models for at least some of said plurality of words;

matching an utterance from one of said users to one of said user-independent word models; and

matching another utterance from said one of said users to one of said user-dependent word models.

7. (amended) [The] A method [according to claim 6 further] of enhancing speech recognition comprising:

providing a set of user-independent word models derived from utterances of a plurality of speakers;

providing a set of user-dependent word models for ones of a plurality of users each derived from utterances of one of said users by:

inviting a new user to speak training words for deriving a set of user-dependent word models;

deriving said set of user-dependent models from said training words; and

storing said set of user-dependent word models;

matching an utterance from one of said users to one of said user-independent word models; and

matching an other utterance from said one of said users to one of said user-dependent word models.

8. (amended) The method according to claim 7 further comprising:

inviting [a] said new user to speak training utterances of a word upon a predetermined number of failures to identify [the] said word among said user-independent word models when no model for said word is present in said user-[in] dependent models;

deriving a word model from said training utterances; and

storing the derived word model in said set of user-dependent word models.

13. (amended) A voice messaging system, comprising a speech recognition system for controlling operation of said voice messaging system, said speech recognition system comprising:

a memory storing:

a first set of word models[,] for recognizing speech independent of [the] an identity of a user, said first set of word models for recognizing a plurality of system commands controlling operation of said voice messaging system; and

a second set of models[,] for recognizing speech of a particular user, at least one model of said second set for initiating performance of at least one of said plurality of system commands, so that at least one of said system commands may be performed in response to a recognized user chosen word.

14. (amended) The voice messaging system of claim 13, wherein said memory further contains computer executable instructions[,] adapting said system to record utterances by said particular user to form said second set, and to collect indicators of system commands to be associated with each model in said second set.

18. (amended) A computer readable medium, storing:

a first set of recognition models[,] for recognizing speech independent of [the] an identity of a user at a speech recognition system, at least some of said models in said first set for recognizing a plurality of system commands; and

computer executable instructions, that when executed at said speech recognition system, adapt said speech recognition system to form and store a second set of models, for recognizing speech of a particular user, with at least one model of said second set for initiating performance of at least one of said plurality of system commands. so that at least one of said system commands may be performed in response to a recognized word chosen by said particular user.

If any issues arise, or if the Examiner has any suggestions for expediting allowance of this Application, the Applicant respectfully invites the Examiner to contact the undersigned at the telephone number indicated below or at ***dvenglarik@davismunck.com***.

The Commissioner is hereby authorized to charge any additional fees connected with this communication or credit any overpayment to Davis Munck Deposit Account No. 50-0208.

Respectfully submitted,

DAVIS MUNCK, P.C.

Date: 2-3-03

Daniel E. Venglarik
Registration No. 39,409

P.O. Drawer 800889
Dallas, Texas 75380
(972) 628-3621 (direct dial)
(972) 628-3600 (main number)
(972) 628-3616 (fax)
E-mail: *dvenglarik@davismunck.com*